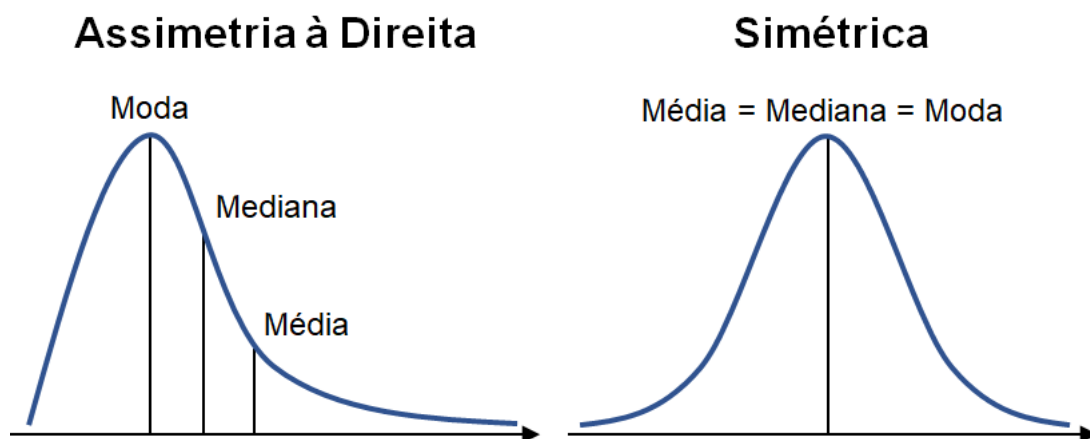


## Transcrição

Anteriormente, abordamos os cálculos das três Medidas de Tendência Central: média, mediana e moda. Neste passo, falaremos sobre a **relação** entre dados.

Por exemplo, poderemos descobrir se será necessário fazer uma transformação nos dados, se a variável é simétrica ou não, e caso seja assimétrica, se

Em nosso notebook, encontraremos a parte "3.4 Relação entre média, mediana e moda" com uma representação gráfica dos formatos padrões mais co



A variável com **Simetria** é um padrão mais difícil de se encontrar, mas é o mais desejado, onde são iguais ou muito próximas e possuem uma distrib

A **Assimetria à Direita** acontece quando há uma grande quantidade de valores que se afunilam de alguma forma, como no caso da pesquisa PNAD s "moda", enquanto a minoria possui grandes rendimentos que diminuem o valor da média.

Basicamente a mesma coisa acontece na **Assimetria à Esquerda**, mas a "força" que "puxa" os dados está na direção contrária.

Com essas representações gráficas, veremos com clareza onde cada estatística descritiva está posicionada.

Nas células seguintes, faremos análises com nossos próprios dados por meio de **gráficos**. Em uma variável `ax`, aplicaremos o método `.distplot()`

Em seguida, *setaremos* seu tamanho com `.figure` e `.set_size_inches()` como 12 e 6 polegadas.

Com isso, geraremos um gráfico chamando `ax` ao final.

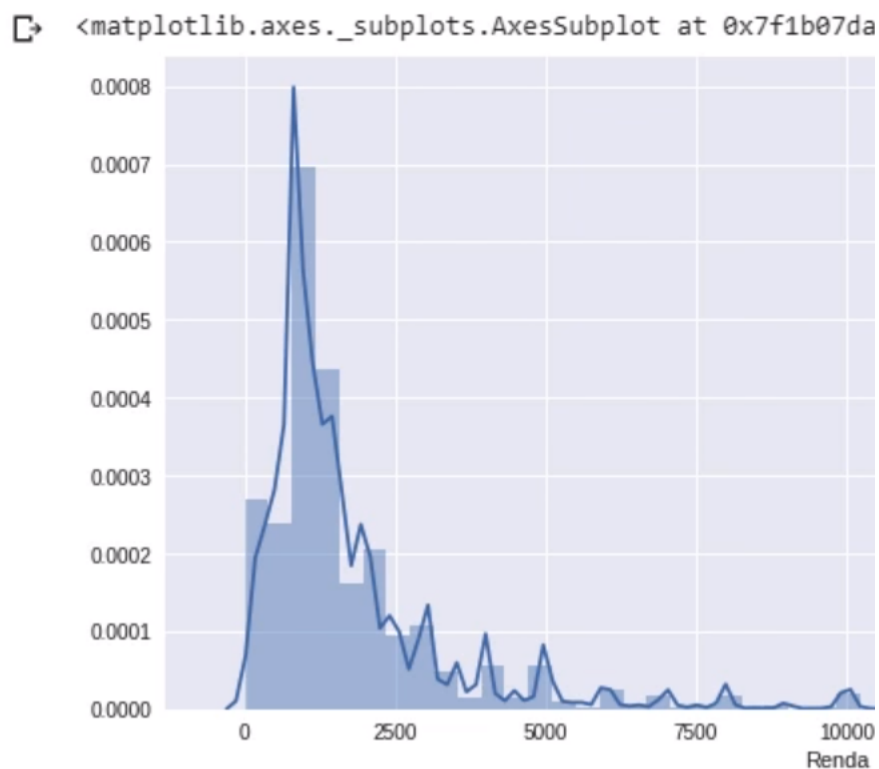
```
ax = sns.distplot(dados.Renda)
ax.figure.set_size_inches(12, 6)
ax
```

Como resultado, teremos um gráfico com Assimetria à direita, conforme já citamos.

Faremos uma modificação para representarmos melhor, pois a desigualdade é tão grande que a leitura do comportamento do gráfico fica um pouco p

Cortaremos os dados adicionando `.query()` após `dados` para selecionarmos as rendas menores que 20000. Veremos essa questão novamente em

```
ax = sns.distplot(dados.query('Renda < 20000').Renda)
ax.figure.set_size_inches(12, 6)
ax
```



Como resultado, teremos um gráfico com uma Simetria à Direita bastante brusca.

Verificaremos as estatísticas para vermos se o comportamento das medidas realmente correspondem ao gráfico.

Na célula seguinte, criaremos a variável `Moda` sendo igual a `dados.Renda.mode()[0]`, e depois a chamaremos.

```
Moda = dados.Renda.mode()[0]
Moda
```

O resultado da moda será `788`, como já vimos.

Calcularemos a Mediana com `median()` e a Média com `mean()` nas células seguintes.

```
Mediana = dados.Renda.median()
Mediana
```

```
Media = dados.Renda.mean()
Media
```

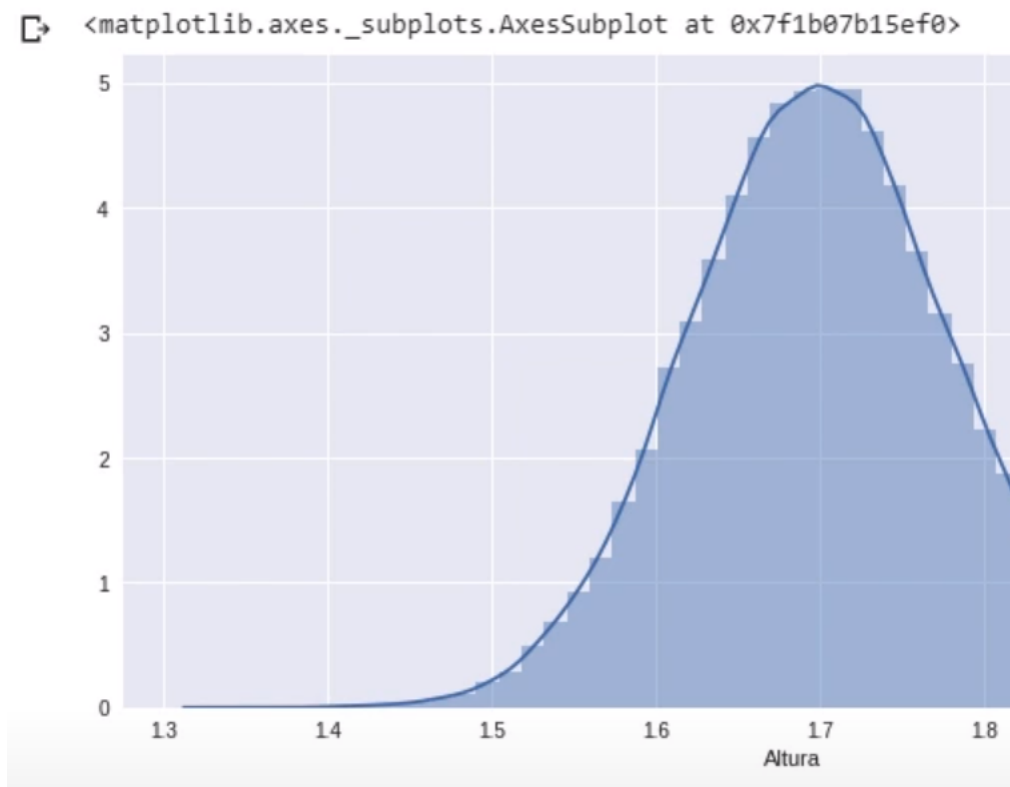
Os retornos serão `1200.0` e `2000.3831988547631` para cada cálculo, correspondendo à mediana e à média respectivamente.

Para confirmarmos se a moda é menor do que a mediana que é menor que a média, escreveremos `Moda < Mediana < Media` na célula seguinte.

Se retornarmos `True`, significará que confirmamos os comportamentos como uma variável simétrica à direita.

Continuando, iremos para a parte de avaliação da variável `Altura`. Construiremos seu gráfico da mesma maneira feita com a anterior.

```
ax = sns.distplot(dados.Altura)
ax = figure.set_size_inches(12, 6)
ax
```



Em seguida, faremos os cálculos das três medidas estatísticas para vermos se realmente se trata de Simetria.

```
Moda = dados.Altura.mode()
Moda
```

```
Mediana = dados.Altura.median()
Mediana
```

```
Media = dados.Altura.mean()  
Media
```

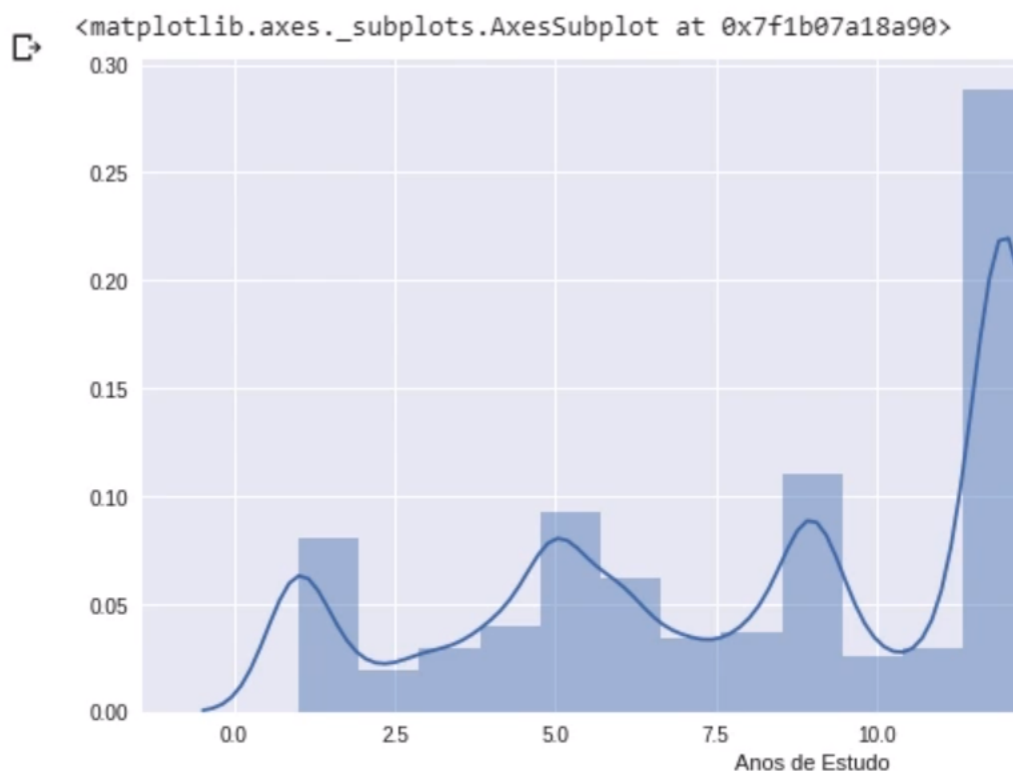
O primeiro resultado representará sua característica multimodal, como já vimos anteriormente. Essa é uma característica bastante forte de variáveis c

O segundo será 1.6993247325 e o terceiro 16995124540575814 . Com isso, veremos que estas duas últimas medidas são bem próximas.

Observando a moda, seus resultados também estarão bem próximos aos dois seguintes. Portanto, poderemos considerar que se trata de Simetria.

Em seguida, aplicaremos os mesmos cálculos para Anos de Estudo com bins sendo igual a 17 .

```
ax = sns.distplot(dados['Anos de Estudo'], bins = 17)  
ax = figure.set_size_inches(12, 6)  
ax
```



O resultado apresentará um gráfico bastante irregular com picos e vales bastante acentuados, e poderemos interpretar de várias formas.

Se considerarmos somente os picos, encontraremos um comportamento bastante parecido com o padrão de Assimetria à Esquerda, pois os maiores p

Em seguida, faremos os cálculos da moda, mediana e média desta variável.

```
Moda = dados['Anos de Estudo'].mode()  
Moda
```

```
Mediana = dados['Anos de Estudo'].median()
```

```
Mediana
```

```
Media = dados['Anos de Estudo'].mean()
```

```
Media
```

O resultado da moda será 12 , a mediana será 11.0 e a média 9.469664237376367 . Comprovaremos novamente ao recebermos True para Moda

Com isso, entenderemos melhor os comportamentos das variáveis e seus gráficos, o que será essencial para nossas análises.

A seguir, abordaremos as **Medidas Separatrizes**.